

REFERENCES

- Ahmad, H., Treude, C., Wagner, M., & Szabo, C. (2024). Smart HPA: A Resource-Efficient Horizontal Pod Auto-Scaler for Microservice Architectures. *2024 IEEE 21st International Conference on Software Architecture (ICSA)*, 46–57. <https://doi.org/10.1109/ICSA59870.2024.00013>
- Camilo, J., & Campos, N. (2024). *Evaluating the Transition from HPA to KEDA in an Event-Driven Architecture Deployed in Kubernetes*.
- Chandrasekhar, A. K., & Chandran, A. S. (2021). *COMPARATIVE ANALYSIS OF LOAD TESTING TOOLS*. <https://opensource-demo.orangehrmlive.com/index.php/dashboard>
- Chilipirea, C., Constantin, A., Popa, D., Crintea, O., & Dobre, C. (2016). Cloud Elasticity: going beyond demand as user load. *Proceedings of the Third International Workshop on Adaptive Resource Management and Scheduling for Cloud Computing, ARMS-CC'16*, 46–51. <https://doi.org/10.1145/2962564.2962570>
- Choonhaklai, P., & Chantrapornchai, C. (2023). Two Autoscaling Approaches on Kubernetes Clusters Against Data Streaming Applications. *2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)*, 1–6. <https://doi.org/10.1109/ITC-CSCC58803.2023.10212432>
- Gao, R., Xie, X., & Guo, Q. (2017). *K-TAHP: A Kubernetes Load Balancing Strategy Base on TOPSIS+AHP*. XX. <https://doi.org/10.1109/ACCESS.2022.Doi>
- HPA Documentation*. (n.d.). Retrieved March 29, 2025, from <https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/>
- Josh Campbell. (n.d.). *Kubernetes vs. Docker: A comprehensive guide to containerization*. Retrieved March 29, 2025, from <https://www.atlassian.com/microservices/microservices-architecture/kubernetes-vs-docker>
- Kubernetes Documentation*. (n.d.). Retrieved February 26, 2025, from <https://kubernetes.io/docs/home/>
- Lorido-Botran, T., Miguel-Alonso, J., & Lozano, J. A. (2014). A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments. *Journal of Grid Computing*, 12(4), 559–592. <https://doi.org/10.1007/s10723-014-9314-7>

- Lukša, M. (n.d.). *Kubernetes in Action*.
- Malallah, H. S., Qashi, R., Abdulrahman, L. M., Omer, M. A., & Yazdeen, A. (2023). Performance Analysis of Enterprise Cloud Computing: A Review. *Journal of Applied Science and Technology Trends*, 4(01), 01–12. <https://doi.org/10.38094/jastt401139>
- Modak, A., Chaudhary, S., Paygude, P., & Ldate, S. (2018). *Techniques to Secure Data on Cloud: Docker Swarm or Kubernetes?* <https://doi.org/10.1109/ICICCT.2018.8473104>
- Molleti, R. (2022). Kubernetes Advanced Auto Scaling Techniques. *Journal of Mathematical & Computer Applications*, 1(4), 1–4. [https://doi.org/10.47363/JMCA/2022\(1\)E126](https://doi.org/10.47363/JMCA/2022(1)E126)
- Nilsen, J. (2023). *Performance Evaluation of Kubernetes Autoscaling strategies on GKE clusters*.
- Prometheus Client libraries*. (n.d.). Retrieved March 5, 2026, from <https://prometheus.io/docs/instrumenting/clientlibs/>
- Prometheus Documentation*. (2026). https://prometheus.io/docs/introduction/first_steps/
- Roman Herbst, N., Kounev, S., & Reussner, R. (n.d.). *Elasticity in Cloud Computing: What It Is, and What It Is Not*.
- Sayfan, G. (2018). *Mastering Kubernetes: Master the art of container management by using the power of Kubernetes*. Packt Publishing Ltd.
- Serrano, N., Gallardo, G., & Hernantes, J. (2015). Infrastructure as a Service and Cloud Technologies. *IEEE Softw.*, 32(2), 30–36. <https://doi.org/10.1109/MS.2015.43>
- Šimon, M., Huraj, L., & Búčik, N. (2023). A Comparative Analysis of High Availability for Linux Container Infrastructures. *Future Internet*, 15(8). <https://doi.org/10.3390/fi15080253>
- Tonini, F., Natalino, C., Wosinska, L., & Monti, P. (2023). Demonstrating the Benefits of Service-Aware Pod Autoscaling with Shared Resources. *2023 IEEE 9th International Conference on Network Softwarization: Boosting Future Networks through Advanced Softwarization, NetSoft 2023 - Proceedings*, 305–307. <https://doi.org/10.1109/NetSoft57336.2023.10175413>
- Wang, T. (2021). *Predictive vertical CPU autoscaling in Kubernetes based on time-series forecasting with Holt-Winters exponential smoothing and long short-term memory*.

What is cloud computing? | Microsoft Azure. (n.d.). Retrieved February 26, 2025, from <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing/>

Wirawan King, R. Z., Hari Trisnawan, P., & Yahya, W. (2024). *Perbandingan Metode Autoscaling Vertical Pod Autoscaler dan Horizontal Pod Autoscaler Kubernetes Pada Google Cloud Platform* (Vol. 1, Number 1). <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13967>

Zhou, H., & Yong, C. H. (2024). Implement HPA for Nginx Service Using Custom Metrics Under Kubernetes Framework. *IEEE Access*, 12, 189722–189734. <https://doi.org/10.1109/ACCESS.2024.3509876>

